ELSEVIER Linear Algebra and its Applications 292 (1999) 139–154

LINEAR ALGEBRA AND ITS APPLICATIONS

Some inequalities for norms on matrices and operators

Jean-Christophe Bourin [1]

Les Côteaux, rue Henri Durel, 78510 Triel, France

Received 24 June 1998; accepted 12 January 1999

Submitted by T. Ando

Abstract

We study several inequalities for norms on matrices, in particular for the Hilbert–Schmidt and operator norms. These inequalites occur when comparing norms of the products XY and YX for matices X and Y with suitable assumptions. We also point out some trace inequalities.

Let $L(H)$ denote the algebra of all bounded linear operators in a separable Hilbert space H. If H has a finite dimension, the only ideals of $L(H)$ are the trivial ones. If $\dim H = \infty$, all proper ideals of $L(H)$ are included in the ideal of compact operators [8, p. 25]. By a unitarily invariant norm $|||\cdot|||$, we mean a norm on an ideal $\mathscr{I}$ of $L(H)$, making $\mathscr{I}$ a Banach space, and such that $|||UXV||| = |||X|||$ for all X in $\mathscr{I}$ and all U, V unitaries in $L(H)$. Examples of unitarily invariant norms are the usual operator norm $\|\cdot\|$ and the Schatten p-norms $(1 \leqslant p < \infty)$, defined for any operator X by

$$\|X\|_p = (\mathrm{Tr}\,|X|^p)^{1/p} = \left(\sum \mu_n^p(X)\right)^{1/p},$$

where $\{\mu_n(X)\}$ are the singular values of X arranged in decreasing order and repeated according to their multiplicities (even if X is not compact, there is a natural definition of $\mu_n(X)$ for all n [4]). The Schatten p-norms act on the $L^p(H)$

[1] E-mail: bourinjc@club-internet.fr

0024-3795/99/$ – see front matter
PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 0 2 4 - 5

ideals, constituted of operators X for which $\|X\|_p < \infty$. The $L^p(H)$ ideals, called Schatten p-classes, are the noncommutative analogues of the classical l^p_n and l^p sequence spaces. More generally, it follows from Schatten's theory [8, chapters 1 and 2], that unitarily invariant norms on ideals are in a one to one correspondance with Banach sequence spaces endowed with symmetric norms.

A noncommutative problem then appears: evaluating differences in the way of regrouping factors in an expression with a product. For instance if the product AB of two operators A and B is normal, we know that [8, p. 95], :

$$\|AB\|_p \leqslant \|BA\|_p.$$

A more recent example [2,5], whose first version was proved by Corach–Porta–Recht is

$$|||STS^{-1} + S^*TS^{*-1}||| \geqslant 2|||T|||,$$

in which S is an invertible operator of $L(H)$ and T is in an ideal with a unitarily invariant norm.

This text focuses on the inequalities linked with these problems of regrouping, and more precisely on inequalities of the following type:

Proposition 1. *Let B be a positive operator, E a projection and Ψ an increasing positive function defined on the spectrum of B. Then*

$$\|BE\Psi(B)\| \leqslant \|EB\Psi(B)\|.$$

We call such an inequality, a gathering inequality. Throughout this paper the term operator with no other precision means a bounded linear operator on H.

1. Positive operators and monotone functions

We say that two positive (i.e. nonnegative) functions defined on a set X have the same (resp. opposite) monotony if

$$f(x) \leqslant f(y) \iff g(x) \leqslant \text{ (resp. } \geqslant \text{) } g(y) \quad \text{for all } x, y \text{ in } X.$$

Let (Ω, P) be a probability space. We denote by

$$E(f) = \int_\Omega f(\omega)\ \mathrm{d}P(\omega),$$

the expectation of an integrable function f.

Lemma 1. *Let f and g be two positive measurable functions on a probability space. If f and g have the same* (*resp. opposite*) *monotony,*

$E(f)E(g) \leqslant$ (*resp.* $\geqslant$) $E(fg)$.

Proof. We prove the case of same monotony. It is very easy to show that for any x and y in the probability space Ω, we have the elementary estimate

$$f(x)g(y) + f(y)g(x) \leqslant f(x)g(x) + f(y)g(y).$$

Then

$$\begin{aligned} E(f)E(g) &= \int_{\Omega\times\Omega} f(x)g(y)\,\mathrm{d}P(x)\,\mathrm{d}P(y) \\ &= \frac{1}{2}\int_{\Omega\times\Omega} [f(x)g(y) + f(y)g(x)]\,\mathrm{d}P(x)\,\mathrm{d}P(y) \\ &\leqslant \frac{1}{2}\int_{\Omega\times\Omega} [f(x)g(x) + f(y)g(y)]\,\mathrm{d}P(x)\,\mathrm{d}P(y) = E(fg). \qquad \square \end{aligned}$$

Proof of Proposition 1. For any $\varepsilon > 0$ there exists $f \in H$ with $\|f\| = 1$ such that $\|BE\Psi(B)\| - \varepsilon \leqslant \|BE\Psi(B)f\|$. Write $E\Psi(B)f = \alpha h$ with $\|h\| = 1$. Then the projection $h \otimes h$ (this notation means $h \otimes h(g) = \langle h, g\rangle h \ \forall g \in H$ – the scalar product is linear in the second variable) satisfies

$$\|BE\Psi(B)\| - \varepsilon \leqslant \|B\, h \otimes h\, \Psi(B)\| = \langle B^2 h, h\rangle^{1/2} \langle \Psi^2(B)h, h\rangle^{1/2}.$$

Since Ψ is increasing, Lemma 1 shows that the last expression is less than

$$\langle B^2\Psi^2(B)h, h\rangle^{1/2} = \|h \otimes h\ B\Psi(B)\| \leqslant \|EB\Psi(B)\|.$$

Letting $\varepsilon \to 0$, we get the proposition. $\square$

We may ask if for Ψ decreasing, Proposition 1 holds with the $\geqslant$ inequality sign. This is not true: for instance, if

$$B = \begin{pmatrix} 1+\varepsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Psi(B) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix},$$

then, with a small ε, $\|BE\Psi(B)\| < \|EB\Psi(B)\|$.

Proposition 2. *Let A be a self-adjoint operator, B a positive operator and Ψ an increasing positive function defined on the spectrum of B. Then:*
1. *If Ψ is increasing,* $\|BA\Psi(B)\|_2 \leqslant \|AB\Psi(B)\|_2$
2. *If Ψ is decreasing,* $\|BA\Psi(B)\|_2 \geqslant \|AB\Psi(B)\|_2$.

Proof. We make the proof for $n \times n$ matrices. Let $B = \sum_{i=1}^n b_i e_i \otimes e_i$ and $A = \sum_{i,j=1}^n a_{i,j} e_i \otimes e_j$. We have

$$\|AB\Psi(B)\|_2^2 = \sum_{i,j} \mid a_{i,j} \mid^2 b_j^2 \Psi(b_j)^2$$

and

$$\|BA\Psi(B)\|_2^2 = \sum_{i,j} b_i^2 \mid a_{i,j} \mid^2 \Psi(b_j)^2,$$

using $a_{i,j} = \overline{a_{j,i}}$, we compute $\|AB\Psi(B)\|_2^2 - \|BA\Psi(B)\|_2^2$:

$$\sum_{i<j} \mid a_{i,j} \mid^2 \left(b_j^2\Psi(b_j)^2 + b_i^2\Psi(b_i)^2 - b_i^2\Psi(b_j)^2 - b_j^2\Psi(b_i)^2\right).$$

Then the estimates $xf(x) + yf(y) \geqslant (\text{resp.} \leqslant) xf(y) + yf(x)$ for nonnegative reals x, y and positive increasing (resp. decreasing) function f, yield the proposition. □

This shows in particular that Proposition 1 is true for the Hilbert–Schmidt norm $\|\cdot\|_2$. Is Proposition 1 still true for any Schatten p-norm?

Like the previous proposition, most results in this paper are expressed in terms of a pair $(B, \Psi(B))$ with Ψ increasing (or decreasing). It is always straightforward to extend the results to a pair $(f(B), g(B))$ with f and g of same monotony (or of opposite monotony).

We say that two positive operators A and B have the same (resp. opposite) monotony if there exist a positive operator C and two positive functions f and g with the same (resp. opposite) monotony, defined on the spectrum of C, such that $A = f(C)$ and $B = g(C)$.

We denote by S_H the unit sphere of H. Using Lemma 1, the following characterization of same/opposite monotony is easy to state (at least when H has a finite dimension).

Proposition 3. *Let A and B be two positive operators. The following conditions are equivalent:*

(i) $\|ABh\| \geqslant$ (*resp.* $\leqslant$) $\|Ah\|.\|Bh\|$ *for all* $h \in \mathrm{S}_H$.

(ii) *A and B have the same* (*resp. opposite*) *monotony.*

Let us give a sketch of the proof for the $\geqslant$ sign when $\dim H = n$ is finite. (ii)⇒(i) follows from Lemma 1. To prove (i)⇒(ii), let $e_1, \dots e_n$ be eigenvectors corresponding to $\mu_1(A), \dots \mu_n(A)$. By inequality (i) applied successively to $h = e_1, \dots e_n$, we can deduce that the eigenspaces of A are invariant for B. Thus A and B commute. It is then easy to show that A and B have same monotony.

2. Gathering inequalities for the Hilbert–Schmidt norm

In this section we study some inequalities for the Hilbert–Schmidt norm which improve Proposition 2.

We begin by introducing the *hyponormality index* of an operator. This is a number which measures the lack of normality of an operator on a finite dimensional space H. If H has an infinite dimension, then this number measures the lack of hyponormality:

We define the hyponormality index of an invertible operator X by

$$v(X) = \|X^*X^{-1}\|.$$

If X is no longer invertible, we set

$$v(X) = \lim_{\varepsilon \to 0} \|X^*(\mid X \mid + \varepsilon)^{-1}\|.$$

Equivalently,

$$v^2(X) = \min\{k \in \mathbb{R}_+ \mid XX^* \leqslant kX^*X\},$$

or again,

$$v(X) = \sup \frac{\|X^*h\|}{\|Xh\|} \quad (\text{and } v(0) = 1),$$

where the supremum runs over all the vectors h such that $\|Xh\| \neq 0$.

If $v(X)$ is finite, we have $XX^* \leqslant v^2(X)X^*X$, so $\|X^*\| \leqslant v(X)\|X\|$. This shows that $v(X) \in [1, \infty]$. Moreover $v(X) = 1$ if and only if X is hyponormal. In particular, if X is compact, then $v(X) = 1$ implies the normality of X. Indeed, it is easy to check that a compact hyponormal operator is normal; more generally Putnam inequality [7] ensures that a hyponormal operator whose spectrum has zero area is normal.

Now we are ready to state the main result of this section in its full generality.

Theorem 1. *Let A be an operator, B a positive operator and Ψ a positive function defined on the spectrum of B.*

(1) *If Ψ is increasing and $v(A)$ is finite,*

$$\|BA\Psi(B)\|_2 \leqslant v(A)\|AB\Psi(B)\|_2.$$

The $v(A)$ constant is optimal. If A is hyponormal, the inequality holds with $v(A) = 1$.

(2) *If Ψ is decreasing, A is normal and if either A is in the Hilbert–Schmidt class or A is self-adjoint or B is compact,*

$$\|BA\Psi(B)\|_2 \geqslant \|AB\Psi(B)\|_2.$$

Proof. 1. Proof of assertion (1).

(1) First, we suppose that B has a finite rank, and we follow two steps.

• If r is a fixed positive real, the function f defined on $]-r,r[$, $f(s)=\|B^{r+s}AB^{r-s}\|_2^2$ is convex. This can be seen in the calculation of the second derivative of $s\to \mathrm{Tr}\,B^{r-s}A^*B^{2(r+s)}AB^{r-s}$, or, more quickly by remarking that if $(e_1,\ldots,e_n)$ is an orthonormal system associated to B's nonzero eigenvalues and if $a_{ij}=\langle e_i,Ae_j\rangle$, then

$$f(s)=\sum_{i,j}b_i^{2(r+s)}|a_{ij}|^2b_j^{2(r-s)},$$

which is obviously convex.

Besides, f can be extended by continuity to r and $-r$. If we call E the projection on ran(B), we have

$$f(-r)=\|EAB^{2r}\|_2^2\leqslant\|AB^{2r}\|_2^2$$

and

$$f(r)=\|B^{2r}AE\|_2^2\leqslant\|B^{2r}A\|_2^2\leqslant v^2(A)\|AB^{2r}\|_2^2.$$

The convexity of f entails $f(s)\leqslant\sup\{f(-r);\ f(r)\}$, hence

$$f(s)\leqslant v^2(A)\|AB^{2r}\|_2^2$$

which can also be written as

$$\|B^sAB^t\|_2\leqslant v(A)\|AB^{s+t}\|_2 \qquad (0\leqslant s,t).$$

• Let us show that for any increasing $\Psi:\mathbb{R}_+\to\mathbb{R}_+$, $\|BA\Psi(B)\|_2\leqslant v(A)\|AB\Psi(B)\|_2$, or,

$$\mathrm{Tr}\,A^*B^2A\Psi^2(B)\leqslant v^2(A)\,\mathrm{Tr}|A|^2B^2\Psi^2(B).$$

Setting $C=B^2$ and $\varphi=\Psi^2\circ\sqrt{\ }$; we have to prove that for any positive operator of finite rank C and any positive, increasing $\varphi:\ \mathbb{R}_+\to\mathbb{R}_+$, we have:

$$\mathrm{Tr}\,A^*CA\varphi(C)\leqslant v^2(A)\mathrm{Tr}\ |A|^2\ C\varphi(C). \tag{1}$$

The set of the functions φ which verify (1) obviously includes the set Φ of the positive, increasing functions which verify (1) and

$$\mathrm{Tr}\,A^*\varphi(C)AC\leqslant v^2(A)\,\mathrm{Tr}\ |A|^2\ C\varphi(C). \tag{2}$$

Let us show that Φ coincides with the set containing all the positive, increasing functions. Φ is preserved under:

(a) a linear combination with positive coefficients,

(b) "dilation": $\varphi\in\Phi\Rightarrow\varphi_\lambda(x)=\varphi(\lambda x)\in\Phi$,

(c) a pointwise limit, and

(d) if $\varphi \in \Phi$ is continuous and strictly increasing, with $\varphi(0) = 0$ and $\varphi(\infty) = \infty$; then the reciprocal function φ^{-1} is also an element of Φ.

(a)–(c) entail that we just have to prove that $\chi_{[1,\infty[} \in \Phi$ to conclude that any positive, increasing function on $\mathbb{R}_+$ is included into Φ. Thanks to our first step, we know that the functions $x \to x^s$ $(s \geqslant 0)$ belong to Φ. So,

$$\varphi_n(x) = \frac{1}{n}x^n + x^{1/n},$$

is an element of Φ. The reciprocal functions φ_n^{-1} pointwise converge to $\chi_{[1,\infty[}$, and the theorem is proved.

(2) Now, B no longer has a finite rank. If B can be diagonalized, there exists an increasing sequence of operators with a finite rank B_n which commute two by two and strongly converge towards B. We have

$$\|B_n A\Psi(B_n)\|_2 \uparrow \|BA\Psi(B)\|_2 \quad \text{and} \quad \|AB_n\Psi(B_n)\|_2 \uparrow \|AB\Psi(B)\|_2$$

which proves the theorem when B can be diagonalized. The general case can be deduced from it, because for any $\varepsilon > 0$ there exists B_ε which can be diagonalized and which commutes with B such that

$$(1-\varepsilon)B \leqslant B_\varepsilon \leqslant (1+\varepsilon)B \quad \text{and} \quad (1-\varepsilon)\Psi(B) \leqslant \Psi(B_\varepsilon) \leqslant (1+\varepsilon)\Psi(B).$$

We still have to check that $v(A)$ is the best constant. Let $\varepsilon > 0$ and let h be a norm-one vector for which

$$\frac{\|A^*h\|}{\|Ah\|} \geqslant v(A) - \varepsilon.$$

We take $B = h \otimes h$ and $\Psi(B) = I$, where I is the identity of $L(H)$. Letting ε tend towards 0, we see that $v(A)$ is the best constant independent of B and Ψ.

2. Proof of assertion (2).

Let us observe that if B is a positive operator, Θ is a positive increasing function defined on the spectrum of B and A is a normal operator, assertion (1) implies the following gathering inequality for the trace norm:

$$\|BA\Theta(B)A^*B\|_1 \leqslant \|AB^2\Theta(B)A^*\|_1.$$

Now to prove assertion (2), we first assume that A is a normal Hilbert–Schmidt operator. Since Ψ decreases, Ψ is bounded and we can write Ψ^2 as the difference between a constant $k = \Psi^2(0)$ and an increasing function Θ: $\Psi^2 = k - \Theta$. So, $\|BA\Psi(B)\|_2^2 = \|BA\Psi^2(B)A^*B\|_1 = \|BA(k-\Theta)A^*B\|_1 = k\|BAA^*B\|_1 - \|BA\Theta(B)A^*B\|_1$, where the assumption that $A \in L^2(H)$ is *essential* for the last expression to be meaningful. Using the normality of A, and the previous gathering trace norm inequality, we can conclude:

$$\|BA\Psi(B)\|_2^2 \geqslant k\|BA^*AB\|_1 - \|AB^2\Theta(B)A^*\|_1 = \|AB\Psi(B)\|_2^2.$$

We assume now that A is self-adjoint. We may suppose that B can be diagonalized, so there exits an increasing sequence $\{E_n\}$ of finite rank projections commuting with B such that

$$\|BA\Psi(B)\|_2 = \lim \|BE_nAE_n\Psi(B)\|_2.$$

By the first step $\|BE_nAE_n\Psi(B)\|_2 \geqslant \|E_nAE_nB\Psi(B)\|_2$ and we deduce the result by letting n tend to the infinite.

Finally we assume that B is compact. Since Ψ is decreasing and B is a positive compact operator, it is easy to see that there is a sequence $\{B_n\}$ of positive Hilbert–Schmidt operators such that $\Psi(B) = \Psi_n(B_n)$, with Ψ_n decreasing, $\|B_nA\Psi(B)\|_2 \uparrow \|BA\Psi(B)\|_2$ and $AB_n\Psi(B) \to AB\Psi(B)$ in Strong Operator Topology. Thanks to the SOT lower semi-continuity of the Hilbert–Schmidt norm,

$$\|AB\Psi(B)\|_2 \leqslant \liminf \|AB_n\Psi(B)\|_2.$$

In this way, it suffices to show the inequality when $B \in L^2(H)$. It is then possible to reproduce the argument of the first step, using this time as an *essential* assumption the fact that $B \in L^2(H)$ instead of $A \in L^2(H)$. $\square$

Corollary 1. *Let A and B be $n \times n$ matrices with A normal and B positive and let Ψ be a positive function defined on the spectrum of B. Then:*

(1) *If Ψ is increasing,* $\|BA\Psi(B)\| \leqslant \sqrt{n}\|AB\Psi(B)\|$

(2) *If Ψ is decreasing,* $\|AB\Psi(B)\| \leqslant \sqrt{n}\|BA\Psi(B)\|$.

Proof. For Ψ increasing (the decreasing case is similar),

$$\|BA\Psi(B)\| \leqslant \|BA\Psi(B)\|_2 \leqslant \|AB\Psi(B)\|_2 \leqslant \sqrt{n}\|AB\Psi(B)\|. \quad \square$$

Proposition 2 and Theorem 1 are no longer true if one substitutes a Schatten p-norm to the Hilbert–Schimdt norm. For instance, let us consider 3×3 matrices:

$$B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1-\varepsilon \end{pmatrix}, \quad \Psi(B) = \begin{pmatrix} 1+\varepsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then for any $p > 2$ one can find a small ε such that $\|BA\Psi(B)\|_p > \|AB\Psi(B)\|_p$. In dimension 4, we have found counterexamples, for some values of p and for A positive. There are similarly counterexamples for Ψ decreasing. It would be desirable to find counterexamples for $1 \leqslant p < 2$.

3. An operator norm inequality

In this section, we give two applications of the previous theorem. Actually the first one is a trace inequality which can be deduced from Proposition 2. We need the Loewner theorem which states that the functions $t \to t^\alpha$, $0 \leqslant \alpha \leqslant 1$, are operator monotone

$$0 \leqslant A \leqslant B \quad \Rightarrow \quad 0 \leqslant A^\alpha \leqslant B^\alpha.$$

We use the notation $X \ll Y$ to mean that X and Y are two positive operators with either $0 \leqslant X \leqslant Y$ or X, Y are invertible and $\log X \leqslant \log Y$. In Section 4 we shall give more information on the $\ll$ sign.

Proposition 4. *Let X, Y be two positive operators such that $X \ll Y$ and M a positive trace class operator which commutes with X. Then, if α and β are two positive reals, we have*

$$\mathrm{Tr}\, MX^\alpha Y^\beta \leqslant \mathrm{Tr}\, MY^{\alpha+\beta}.$$

Proof. First, we prove the case $X \leqslant Y$. We assume that $\alpha \leqslant 1$. By Loewner's theorem, $X^\alpha \leqslant Y^\alpha$, so

$$\begin{aligned}\mathrm{Tr}\, MX^\alpha Y^\beta &= \|Y^{\beta/2}M^{1/2}X^{\alpha/2}\|_2^2 \\ &\leqslant \|Y^{\beta/2}M^{1/2}Y^{\alpha/2}\|_2^2 \\ &\leqslant \|M^{1/2}Y^{\beta/2+\alpha/2}\|_2^2 \quad \text{(by gathering)} \\ &= \mathrm{Tr}\, MY^{\alpha+\beta}.\end{aligned}$$

The case $\alpha \geqslant 1$ can be deduced by repeating the process.

Now, we assume that X and Y are invertible and $\log X \leqslant \log Y$. By homogeneity, we may assume that $I \leqslant X, Y$ (where I denotes the identity operator) , so $0 \leqslant \log X \leqslant \log Y$, and

$$\mathrm{Tr}\, MX^\alpha Y^\beta = \mathrm{Tr}\, M\left(\sum_n \frac{(\alpha \log X)^n}{n!}\right)\left(\sum_k \frac{(\beta \log Y)^k}{k!}\right)$$

$$\mathrm{Tr}\, MX^\alpha Y^\beta \leqslant \sum_{n,k} \frac{1}{n!k!} \mathrm{Tr}\, M(\alpha \log Y)^n (\beta \log Y)^k = \mathrm{Tr}\, MY^{\alpha+\beta}. \qquad \square$$

Remark. Proposition 4 gives an immediate proof of the McCarthy inequality (cf. [5], p. 20, Theorem 1.22)

$$\mathrm{Tr}(X+Y)^p \geqslant \mathrm{Tr} X^p + \mathrm{Tr} Y^p \quad (0 \leqslant X,\ Y;\ p \geqslant 1).$$

Indeed, $\mathrm{Tr}(X+Y)^p = \mathrm{Tr}X(X+Y)^{p-1} + \mathrm{Tr}\,Y(X+Y)^{p-1} \geqslant \mathrm{Tr}X^p + \mathrm{Tr}\,Y^p$. Similarly we also get

$$\mathrm{Tr}(X+Y)^p \geqslant \mathrm{Tr}X^p + \mathrm{Tr}\,Y^p + \mathrm{Tr}\,(XY^{p-1} + YX^{p-1}) \quad (0 \leqslant X,\, Y;\; p \geqslant 2).$$

Proposition 4 is not necessarily true when β is negative. For instance taking $\alpha = 3,\ \beta = -7$ and

$$M = X = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}, \qquad Y = \begin{pmatrix} 9 & 0 \\ 0 & 2 \end{pmatrix}$$

one has $\mathrm{Tr}X^4Y^{-7} > \mathrm{Tr}XY^{-4}$. However, we have the following result:

Proposition 4a. *Let X, Y be two positive operators with Y invertible and $X \leqslant Y$. If M is a positive trace class operator which commutes with X and α, β are two reals such that $-1 \leqslant \beta$ and $0 \leqslant \alpha + \beta$, we have*

$$\mathrm{Tr}MX^\alpha Y^\beta \leqslant \mathrm{Tr}MY^{\alpha+\beta}.$$

Proof. We have just to prove the case $-1 \leqslant \beta < 0$. By repeating the process, we may assume that $\alpha \leqslant 1$. By a limit argument, we may assume that X is invertible. There exist an orthonormal system $\{e_n\}$ and two sequences of reals $\{x_n\}$ and $\{m_n\}$ such that $M = \sum_n m_n e_n \otimes e_n$ and $X(e_n) = x_n e_n$. Thus

$$\mathrm{Tr}MY^{\alpha+\beta} = \sum_n m_n x_n^\alpha \langle e_n, X^{-\alpha} e_n\rangle \langle e_n, Y^{\alpha+\beta} e_n\rangle.$$

Since $t \to t^{-\alpha}$ is operator decreasing,

$$\mathrm{Tr}MY^{\alpha+\beta} \geqslant \sum_n m_n x_n^\alpha \langle e_n, Y^{-\alpha} e_n\rangle \langle e_n, Y^{\alpha+\beta} e_n\rangle.$$

Since $t \to t^{-\alpha}$ decreases and $t \to t^{\alpha+\beta}$ increases, Lemma 1 implies

$$\mathrm{Tr}MY^{\alpha+\beta} \geqslant \sum_n m_n x_n^\alpha \langle e_n, Y^\beta e_n\rangle = \mathrm{Tr}MX^\alpha Y^\beta. \quad \square$$

A more original application of the previous theorem is an operator norm inequality. We say that a normal operator X is *semi-unitary* if its restriction to $\mathrm{ran}(X)$ is a unitary operator.

Theorem 2. *Let B be a positive operator, E a semi-unitary operator and Ψ an increasing positive function defined on the spectrum of B. Then*

$$\|BE\Psi(B)\| \leqslant \sqrt{2}\|EB\Psi(B)\|.$$

$\sqrt{2}$ is in general the best constant possible; however if E is a projection,

$$\|BE\Psi(B)\| \leqslant \|EB\Psi(B)\|.$$

Proof. The second assertion is just Proposition 1 and we just have to prove the first assertion. By a limit argument, we may assume that there is $h \in H$, $\|h\| = 1$, such that

$$\|BE\Psi(B)\| = \|BE\Psi(B)h\|.$$

Let P be the projection E^*E and set $f = \Psi(B)h$. Since $\|Pf\| = \|Ef\|$ we obtain the semi-unitary operator R of rank 2 such that $\mathrm{ran}(R) = \mathrm{Span}\{Pf, EPf\}$ and $RPf = Ef$. Since $R^*R \leqslant P$ we have $Ef = Rf$ and

$$\begin{aligned} \|BE\Psi(B)\| = \|BEf\| &= \|BRf\| \\ &\leqslant \|BR\Psi(B)\| \\ &\leqslant \|BR\Psi(B)\|_2 \\ &\leqslant \|RB\Psi(B)\|_2 \quad \text{(by gathering)} \\ &\leqslant \sqrt{2}\|RB\Psi(B)\| \quad (\mathrm{rank}(R) = 2) \\ &\leqslant \sqrt{2}\|EB\Psi(B)\| \quad (R^*R \leqslant E^*E). \end{aligned}$$

To see that the constant $\sqrt{2}$ can not be improved, we consider

$$B_n = \begin{pmatrix} \frac{n}{n+1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & n \end{pmatrix}, \quad \Psi(B_n) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_n = \frac{1}{\sqrt{n^2+1}} \begin{pmatrix} 0 & n & 0 \\ n & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then

$$\lim_{n\to\infty} \frac{\|B_n E_n \Psi(B_n)\|}{\|E_n B_n \Psi(B_n)\|} = \sqrt{2},$$

thus the constant $\sqrt{2}$ can not be improved. □

Question. As the author does not have any counterexamples, a question is raised: Is Theorem 2 still valid for any normal operator E ?

4. Gathering inequalities for unitarily invariant norms

Besides the Hilbert–Schmidt and operator norms results of the Section 3, it is natural to study unitarily invariant norms. The reader is referred to Ando's paper [1] for many majorization relations between eigenvalues and the corresponding norm inequalities. In all the sequel, $|||\cdot|||$ denotes a unitarily invariant norm on an ideal $\mathscr{I}$.

Lemma 2. *For an operator $A \in \mathscr{I}$ and two invertible operators B and C, the map $t \to ||| \, |B|^t A \mid C \mid^t \, |||$ is log-convex, equivalently*

$$|||A||| \leqslant |||B^* A C^*|||^{1/2} |||B^{-1} A C^{-1}|||^{1/2}$$

Proof. By unitary invariance, the analytic map $f(z) = \mid B \mid^z A \mid C \mid^z$ satisfies $|||f(x + \mathrm{i}y)||| = |||f(x)|||$ for all reals x and y. Hence the lemma is a straightforward application of the Banach space valued version of the Three lines theorem.

It is possible, rather than the previous interpolation argument, to give a proof based on standard majorization techniques. Note that

$$\begin{aligned} \|A\|^2 = \rho(A^*A) = \rho(CA^*AC^{-1}) &\leqslant \|CA^*AC^{-1}\| \\ &= \|CA^*BB^{-1}AC^{-1}\| \\ &\leqslant \|B^*AC^*\| \, \|B^{-1}AC^{-1}\|, \end{aligned}$$

by using the simple fact that the spectral radius of an operator X is less than or equal to its norm, with equality for X self-adjoint. Thus the inequality is proved in case of the operator norm. An antisymetric tensor product then shows that, if $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ denote the sequences of the respective singular values of A, B^*AC^*, $B^{-1}AC^{-1}$ arranged in decreasing order and repeated according to their multiplicity, we have

$$\prod_{n=1}^{N} a_n \leqslant \prod_{n=1}^{N} b_n^{1/2} c_n^{1/2} \quad \text{for each integer } N.$$

This implies that $\{a_n\}$ is weakly majorized by $\{b_n^{1/2} c_n^{1/2}\}$, so we have

$$\begin{aligned} |||A||| &= \Phi(a_1, a_2, \ldots) \\ &\leqslant \Phi(b_1^{1/2} c_1^{1/2}, b_2^{1/2} c_2^{1/2}, \ldots) \\ &\leqslant \Phi^{1/2}(b_1, b_2, \ldots) \Phi^{1/2}(c_1, c_2, \ldots) \end{aligned}$$

by the Cauchy–Shwartz inequality for the symmetric gauge function Φ corresponding to $||| \cdot |||$ (cf. [3], p. 87). □

Proposition 5. *Let $A \in \mathscr{I}$ be a self-adjoint operator and X, Y, Z three positive operators such that $X^2 = YZ$. Then*

$$|||XAX||| \leqslant |||YAZ|||.$$

Proof. By continuity, we may assume that X, Y, Z are invertible. Then, by Lemma 2,

$$\begin{aligned}&|||XAX|||\\&\quad\leqslant |||Y^{1/2}Z^{-1/2}(XAX)Y^{-1/2}Z^{1/2}|||^{1/2}|||Y^{-1/2}Z^{1/2}(XAX)Y^{1/2}Z^{-1/2}|||^{1/2}\\&\quad= |||YAZ|||^{1/2}|||ZAY|||^{1/2} = |||YAZ|||. \qquad \square\end{aligned}$$

Proposition 6. *Let A be an operator in $\mathscr{I}$ and B a positive operator. Then, for all $s, t \geqslant 0$, we have*

$$|||B^sAB^t||| \leqslant v(A)^{s/(s+t)}|||AB^{s+t}|||.$$

Moreover, if B is invertible and $0 \leqslant s < t$,

$$|||B^sAB^{-t}||| \geqslant v(A)^{s/(s-t)}|||AB^{s-t}|||.$$

Equivalently, if B is invertible and $0 \leqslant t < s$,

$$|||B^sAB^{-t}||| \geqslant v(A^*)^{s/(t-s)}|||AB^{s-t}|||.$$

In the first inequality of the proposition, $|||AB^{s+t}||| = 0 \Rightarrow AB^{s+t} = 0 \Rightarrow B^sAB^t = 0 \Rightarrow |||B^sAB^t||| = 0$. Hence we may adopt the convention that $v(A)0 = 0$ when $v(A) = \infty$.

Proof. By Lemma 2, $f(r) = |||B^{s-r}AB^{t+r}|||$ is log-convex on $]-t, s[$. From the lower semi-continuity of $|||\cdot|||$ in WOT we easily deduce that f can be extended by continuity to $-t$ and s with $f(-t) = |||B^{s+t}AE|||$ and $f(s) = |||EAB^{s+t}|||$ where $E = B^0$ is the support projection of B. Hence,

$$f(0) \leqslant f(-t)^{s/(s+t)}f(s)^{t/(s+t)},$$

where

$$f(0) = |||B^sAB^t|||, \qquad f(-t) \leqslant |||B^{s+t}A|||, \qquad f(s) \leqslant |||AB^{s+t}|||.$$

Also as $|||B^{s+t}A||| \leqslant v(A)|||AB^{s+t}|||$, we get the first inequality.

To prove the second inequality, we consider the function $f(r) = |||B^{s-r}AB^{r-t}|||$. Since $g(r) = \log f(r)$ is convex, the graphic representation of g shows us that the point $(0, g(0))$ is above the line passing by $(s, g(s))$ and $(t, g(t))$. Hence,

$$g(0) \geqslant g(s) + \frac{g(t) - g(s)}{t - s}.(0 - s),$$

thus

$$\log f(0) \geqslant \log f(s) + \frac{-s}{t - s}\log\frac{f(t)}{f(s)},$$

or

$$\log f(0) \geqslant \log \left(f(s)^{t/(t-s)} f(t)^{-s/(t-s)} \right),$$

so

$$f(0) \geqslant f(s)^{t/(t-s)} f(t)^{-s/(t-s)}.$$

Then, using $f(0) = |||B^s AB^{-t}|||$ and $f(t) = |||B^{s-t}A||| \leqslant v(A)|||AB^{s-t}||| = v(A)f(s)$, we deduce the result. The proof of the third inequality is similar. Actually it is not difficult to see that the second and third inequalities are equivalent. $\square$

Remark. In the two previous propositions, it has been assumed that some of the operators involved belong to $\mathscr{I}$. These assumptions are not necessary if $\mathscr{I}$ have the following property: for any sequence $\{X_n\}$ in $\mathscr{I}$ such that $\sup|||X_n||| < \infty$ and $\{X_n\}$ converges weakly to X, then $X \in \mathscr{I}$ and $|||X||| \leqslant \liminf |||X_n|||$. Actually, it follows from Schatten's theory and the Goh'berg–Krein's non-commutative Fatou lemma (cf. [8], Theorem 2.7 pp. 28–29) that $\mathscr{I}$ can be embedded in a unique larger ideal having the previous property. When $\mathscr{I}$ has this property, we say that $|||\cdot|||$ is maximal. It then becomes natural to set $|||A||| = \infty$ if A is not in $\mathscr{I}$ and, in this framework, it should be observed that an inequality such that $|||A||| \leqslant |||B|||$ yields as a corollary $B \in \mathscr{I} \Rightarrow A \in \mathscr{I}$.

The next result is an improvement of the trace inequality of section 3. First of all, we have to give more precisions about the $\ll$ sign. The order relation $\log X \leqslant \log Y$ on positive invertible operators is called the chaotic order and is denoted $\ll$. Since log is operator monotone, the chaotic order is a weaker order than the usual one. Fujii et al. [6] have given a very simple proof of the following result for two invertible, positive operators X, Y

$$\log X < \log Y \qquad \text{iff} \qquad X^r < Y^r \text{ holds for an } r > 0.$$

If X and Y are positive operators with X *not* invertible, we may keep the notation $X \ll Y$ to mean that there exists $\eta > 0$ such that $\log(X + \varepsilon) \leqslant \log(Y + \varepsilon)$ when $0 < \varepsilon \leqslant \eta$.

Proposition 7. *Let X, Y, M be positive operators with $M \in \mathscr{I}$ and α, β two positive reals. Assume that $X \ll Y$ and $MX = XM$. Then,*

$$|||MX^\alpha Y^\beta||| \leqslant |||MY^{\alpha+\beta}|||.$$

Moreover, if $|||\cdot|||$ is maximal, the inequality still holds when $M \notin \mathscr{I}$, and M may even be taken unbounded. (The author does not know if the second assertion of the proposition remains true when $|||\cdot|||$ is not maximal.)

Proof. (1) $M \in \mathscr{I}$. Then

$$|||MX^{\alpha}Y^{\beta}||| = \lim_{\varepsilon \to 0} |||M(X+\varepsilon)^{\alpha}(Y+2\varepsilon)^{\beta}|||$$

and

$$|||MY^{\alpha+\beta}||| = \lim_{\varepsilon \to 0} |||M(Y+2\varepsilon)^{\alpha+\beta}|||.$$

Since $X \ll Y$, we have $\log(X+\varepsilon) < \log(Y+2\varepsilon)$ for ε sufficiently small. The Fujii–Furuta–Kamei's characterization of the chaotic order implies that $(X+\varepsilon)^r < (Y+2\varepsilon)^r$ for an $r>0$. The same process as in the first part of the proof of Proposition 4 then shows that

$$|||M(X+\varepsilon)^{\alpha}(Y+2\varepsilon)^{\beta}||| \leqslant |||M(Y+2\varepsilon)^{\alpha+\beta}|||.$$

Letting $\varepsilon \to 0$, we get the result.

(2) $M \notin \mathscr{I}$ may be unbounded, $|||\cdot|||$ is maximal.

We may assume that $MY^{\alpha+\beta}$ – with a priori domain $\{h \in H \mid Y^{\alpha+\beta}h \in \operatorname{dom} M\}$ – can be extended as an operator in $\mathscr{I}$, otherwise $|||MY^{\alpha+\beta}||| = \infty$ and there is nothing to prove. Therefore we can choose an increasing sequence $\{M_n\}$ of bounded, positive operators, commuting with X and M, such that $M_nY^{\alpha+\beta} \to MY^{\alpha+\beta}$ in Weak Operator Topology. The lower semi-continuity of $|||\cdot|||$ in WOT implies

$$|||MY^{\alpha+\beta}||| = \lim |||M_nY^{\alpha+\beta}|||$$

By case 1, $|||M_nY^{\alpha+\beta}||| \geqslant |||M_nX^{\alpha}Y^{\beta}|||$. Hence $\sup_n\{|||M_nX^{\alpha}Y^{\beta}|||\} < \infty$. Since $\{M_n\}$ increases, $M_nX^{\alpha}Y^{\beta}$ converges in WOT to Z, the bounded closure of $MX^{\alpha}Y^{\beta}$ –whose a priori domain is $\{h \in H \mid Y^{\beta}h \in \operatorname{dom} MX^{\alpha}\}$. Since $|||\cdot|||$ is maximal, $|||MX^{\alpha}Y^{\beta}||| \leqslant \liminf |||M_nX^{\alpha}Y^{\beta}||| \leqslant |||MY^{\alpha+\beta}|||$. $\square$

Proposition 7 clearly appears as a generalization of Proposition 4 (the trace inequality can be rephrased as an inequality for the Hilbert–Schmidt norm). Is it possibe, in a similar way, to generalize Proposition 4a to unitarily invariant norms ?

Acknowledgements

The author is grateful to Frank Hansen and Georges Skandalis for their useful remarks and their friendly support. The author is also indebted to the referee for his comments, particularly for simplified proofs and for having pointed out that inequalities previously written for the Schatten p-norms could be stated in the unitarily invariant norms framework.

References

[1] T. Ando, Majorizations and inequalities in matrix theory, Linear Algebra Appl. 199 (1994) 17–67.
[2] E. Andruchow, G. Corach, D. Stojanoff, Geometric operator inequalities, Linear Algebra Appl. 258 (1997) 295–310.
[3] R. Bhatia (1996), Matrix Analysis, Springer, Germany.
[4] T. Fack, H. Kosaki, Generalized s-numbers of τ - measurable operators, Pacific J. Math. 123 (1986) 269–300.
[5] J.I. Fujii, M. Fujii, T. Furuta, R. Nakamato, Norm inequalities equivalent to Heinz inequality, Proc. Amer. Math. Soc. 118 (1993) 827–830.
[6] M. Fujii, J.F. Jiang, E. Kamei, Characterization of chaotic order and its applications to Furuta's type operator inequalities, Linear Multilinear Algebra 43 (1998) 339–349.
[7] C.R. Putnam, An inequality for the area hyponormal spectra, Math. Z. 116 (1970) 323–330.
[8] B. Simon, Trace Ideals and Their Applications LMS lecture note, 35 Cambridge Univ. Press, Cambridge, 1979.